MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 2: 1–11
DOI 10.3897/mbmg.2.24262

Research Article

Estimating fish alpha- and beta-diversity along a small stream with environmental DNA metabarcoding

Yiyuan Li[1], **Nathan T. Evans**[2], **Mark A. Renshaw**[3], **Christopher L. Jerde**[4], **Brett P. Olds**[3], **Arial J. Shogren**[1,5], **Kristy Deiner**[6], **David M. Lodge**[6], **Gary A. Lamberti**[1], **Michael E. Pfrender**[1,5]

1 *Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA.*
2 *Southeast Environmental Research Center, Florida International University, Miami, FL, USA.*
3 *Oceanic Institute, Hawai'i Pacific University, Waimanalo, HI, USA.*
4 *Marine Science Institute, University of California, Santa Barbara, Santa Barbara, CA, USA.*
5 *Environmental Change Initiative, University of Notre Dame, Notre Dame, IN, USA.*
6 *Department of Ecology and Evolutionary Biology, Atkinson Center for a Sustainable Future, Cornell University, Ithaca, NY, USA.*

Corresponding author: *Yiyuan Li* (yli19@nd.edu)

Academic editor: *M.T. Monaghan* | **Received** 8 February 2018 | **Accepted** 5 April 2018 | **Published** 14 May 2018

Abstract

Environmental DNA (eDNA) metabarcoding has been increasingly applied to biodiversity surveys in stream ecosystems. In stream networks, the accuracy of eDNA-based biodiversity assessment depends on whether the upstream eDNA influx affects downstream detection. Biodiversity assessment in low-discharge streams should be less influenced by eDNA transport than in high-discharge streams. We estimated α- and β-diversity of the fish community from eDNA samples collected in a small Michigan (USA) stream from its headwaters to its confluence with a larger river. We found that α-diversity increased from upstream to downstream and, as predicted, we found a significant positive correlation between β-diversity and physical distance (stream length) between locations indicating species turnover along the longitudinal stream gradient. Sample replicates and different genetic markers showed similar species composition, supporting the consistency of the eDNA metabarcoding approach to estimate α- and β-diversity of fishes in low-discharge streams.

Key Words

eDNA, biodiversity assessment, biomonitoring, α-diversity, β-diversity, high throughput sequencing

Introduction

Freshwater fishes provide vital ecosystem services to humans including protein, recreation (McIntyre et al. 2016) and food web regulation (Holmlund and Hammer 1999). Due to the sensitivity of fishes to environmental stressors, such as chemical pollutants, pH and temperature, fish diversity is often used as a metric of river ecosystem health (Bae et al. 2014; Marzin et al. 2012; Simon and Evans 2017; Vörösmarty et al. 2010).

Fish diversity in rivers is traditionally assessed using a variety of census methods including netting, trapping and electrofishing. The accuracy of these methods relies heavily on the intensity of sampling efforts and the susceptibility of species and individuals to capture (Bonar et al. 2009; Evans et al. 2016). As a result of low sampling efficiencies, traditional sampling methods often underestimate true fish species diversity (Evans and Lamberti 2017; Olds et al. 2016) leading to biased metrics of ecosystem health.

An alternative approach to traditional capture-based survey methods is the use of environmental DNA (eDNA) metabarcoding to estimate fish biodiversity. Environmental DNA is the DNA that organisms shed into the environment through mechanisms such as sloughed tissue, faeces, exudates and decay of carcasses (Barnes and Turner 2016; Ficetola et al. 2008). As only water

Copyright *Yiyuan Li et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PENSOFT

samples are needed for collecting eDNA, this approach is time-efficient with low environmental and organismal impacts and can provide information on species not easily assessed with traditional methods (Deiner et al. 2017; Evans et al. 2017; Jerde et al. 2011; Lodge et al. 2012). Combined with high-throughput sequencing (HTS) technology, eDNA assays can be a cost effective alternative (McInerney and Rees 2018) facilitating biodiversity surveys of diverse environments (Bohmann et al. 2014; Deiner et al. 2017; Thomsen et al. 2012).

In eDNA metabarcoding studies, α-diversity (i.e. the number of species within a defined location) and β-diversity (i.e. the species turnover between locations) are often used to quantify biodiversity patterns. Due to the longitudinal nature of river ecosystems and the flow-dependent movement of eDNA, a reliable estimation of α- and β-diversity in rivers requires an understanding of the eDNA transport along the longitudinal stream gradient of the catchment. In a lotic environment, the eDNA transport distance can vary with discharge (Q, volume water per time), rate of release of DNA into the environment, characteristics of eDNA, entrapment of eDNA particles within benthic substrates and biofilms (Jerde et al. 2016; Shogren et al. 2016, 2017) and environment-specific degradation rates (Barnes and Turner 2016; Seymour et al. 2018). Previous research using standard PCR and qPCR suggest that the DNA of upstream species can often be detected far downstream of the source populations, with the transport distance being largely a function of stream size and discharge. For example, in high-discharge systems (with Q > 3000 L/s), eDNA has been found up to 12 km downstream (Deiner and Altermatt 2014). In contrast, smaller systems with lower discharge (Jerde et al. 2016; Pilliod et al. 2014; Wilcox et al. 2016) have much shorter (5-1459 m) average transport distances. While standard PCR and qPCR typically target individual species one at a time, eDNA metabarcoding can detect an entire multispecies community. Consistent with PCR and qPCR studies, eDNA metabarcoding studies found that eDNA from upstream can be detected 2-10 km downstream (Civade et al. 2016; Deiner et al. 2017). In high-discharge contexts eDNA transport could inflate downstream α-diversity and reduce β-diversity between sites by homogenising the detected diversity amongst sites. An example of this transport effect was shown by Deiner et al. (2016), who found no correlation between β-diversity and longitudinal distance between samples in a river with Q > 3000 L/s. However, the relationship between β-diversity and longitudinal distance in small lotic environments with lower discharge is largely unknown.

To test the relationship between β-diversity and longitudinal distance and the utility of eDNA to describe α- and β-diversity in systems with relatively low discharge, we surveyed freshwater fish diversity using eDNA along the course of a small stream in Michigan, USA from its headwaters to its confluence with a larger river. We hypothesised that β-diversity between sampling sites would correlate with the longitudinal distance between sites.

Materials and methods

Sampling, DNA extraction and library preparation

We collected water samples for eDNA isolation from Eagle Creek, a small first-order stream (Q = 0.1-324 L/s) in the Kalamazoo River watershed that flows approximately 12 km from its source in the Fort Custer Training Center (Michigan Army National Guard) to its confluence with the Kalamazoo River (42.3374N, 85.3377W) near the town of Augusta, Michigan. We sampled six locations in Eagle Creek and two sites in the Kalamazoo River (Figure 1). In the Kalamazoo river, we sampled one site upstream (Location 7) and one site downstream (Location 8) of the confluence of Eagle Creek to detect the new species that Eagle Creek contributed to the river and novel species carried by the Kalamazoo River.

Samples were collected on 29 April 2015, between 0940 and 1230 h. We sampled from the furthest downstream site (Location 8) to the most upstream site (Location 1) (Figure 1). At each sampling location, we collected three 250-ml replicate water samples to yield a total of 24 samples. To avoid contamination, we collected each surface water sample with an extendable pole with a dip bucket attached to the end. To maintain semi-sterile conditions, the pole and dip bucket were wiped with 10% bleach and rinsed with reverse-osmosis (RO) water prior to collecting each additional sample and between locations. For each sampling location, a 250-ml bottle filled with RO water in the laboratory prior to sampling, was transported alongside field sampling bottles to serve as a field control. One field control bottle was opened briefly at each location and resealed in the field. Sample bottles were tightly sealed, wiped with a 10% bleach solution and immediately placed in a cooler containing wet ice for transport to the laboratory. Before being opened and filtered, sample bottles were rinsed with deionised water as in our previous studies (Olds et al. 2016, Evans et al. 2017).

After collecting the water samples, we recorded the latitude and longitude coordinates for each sampling location using a handheld global positioning system (Garmin, Olathe, Kansas, USA). We then measured turbidity, dissolved oxygen (D.O.), pH, water temperature and flow velocity at each site using the cross-sectional approach (Gore and Banning 2017). Surface water turbidity was quantified using a portable turbidity meter (Hach, Loveland, Colorado, USA). Dissolved oxygen, pH and water temperature were quantified using a multi-parameter probe (Yellow Springs Instrument, Ohio, USA). The flow velocity at each site was measured at both the left and right edge of the wetted channel and the thalweg using a Marsh-McBirney Flo-mate flow meter and wading rod (Hach, Loveland, Colorado, USA). Discharge was calculated as mean velocity × width × depth for each sampling site. The distance between the sampling locations was calculated as the linear stream distance between the locations following the stream channel midpoint.

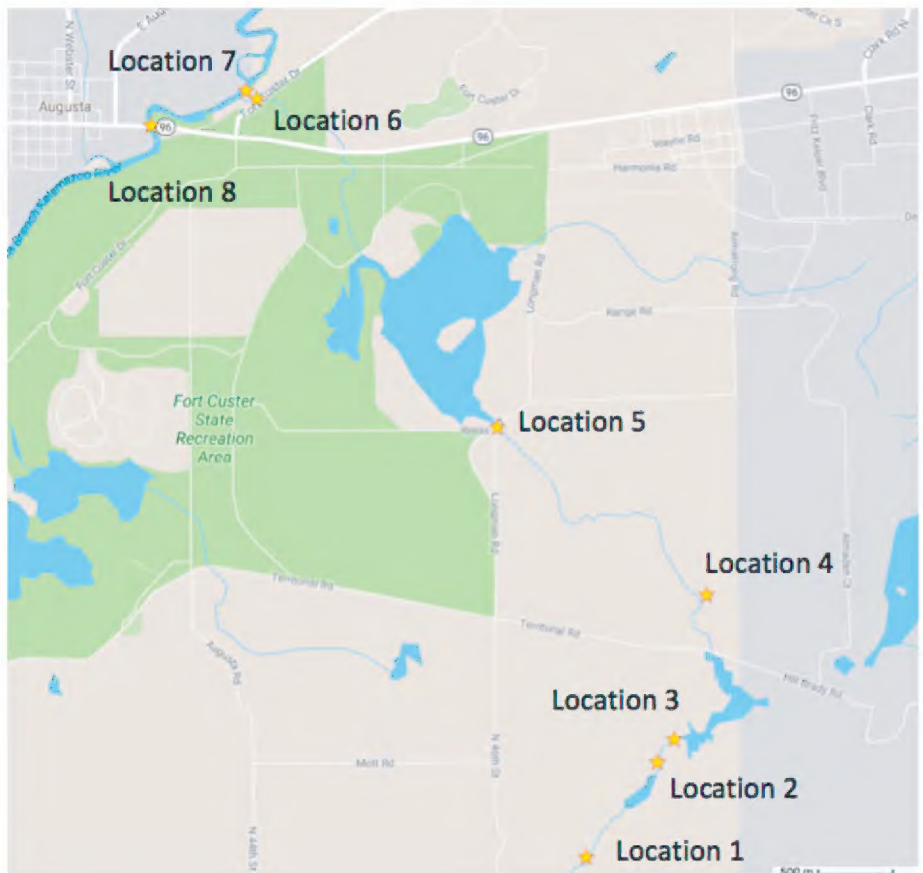

Figure 1. Map of sampling locations along Eagle Creek. Location 1 is the most upstream sampling location in Eagle Creek. Location 6 is the most downstream sampling location in Eagle Creek. Locations 7 and 8 are in the Kalamazoo River. Location 7 is upstream of the confluence of Eagle Creek whereas Location 8 is downstream of the confluence.

Sample processing and extraction

In the laboratory within 6 hours of sample collection, water samples were vacuum-filtered on to 47 mm, 1.2-µm pore size polycarbonate membrane filters (EMD Millipore, Billerica, Massachusetts, USA). The filters were placed in 2.0 ml microcentrifuge tubes containing 700 µl of CTAB and stored at -20 °C as described by Evans et al. (2017). DNA was isolated following a modified Chloroform-Isoamyl alcohol (24:1, Amresco, Dallas, Texas, USA) extraction and an isopropanol precipitation (Olds et al. 2016; Renshaw et al. 2015). To remove potential PCR inhibitors, re-suspended DNA was treated with the OneStep™ PCR Inhibitor Removal Kit (Zymo Research, Irvine, California, USA).

To check for and quantify contamination during the experimental process, we included a series of field controls, mock communities and PCR negative controls (Suppl. material 1). Eight field controls (one for each sampling location) were extracted separately from the eDNA samples and mock community samples, amplified and visualised on agarose gel to detect potential contamination in the field. Four mock community samples were constructed independently and processed along with eDNA samples from DNA extraction to sequencing to detect any potential contamination during the experiment. To avoid the mock community species affecting the results, the mock community samples comprised 60 ng (10 ng per species) of tissue-derived DNA (measured with Qubit) from six Indo-Pacific marine fishes: *Amphiprion ocellaris, Salarias fasciatus, Ecsenius bicolor, Centropyge bispinosa, Pseudanthias dispar* and *Macropharyngodon negrosen-sis*. Mock communities 1 and 2 were processed along with samples from Locations 1-4. Mock communities 3 and 4 were processed along with samples from Locations 5-8. A no-template PCR negative control (sterile water) was used for each of the three gene regions amplified.

Library preparation and sequencing

The 3 replicates from each sampling location (24 samples in total), 4 mock communities and a single PCR negative control were PCR amplified for 3 mitochondrial gene regions. The three PCR primer sets were used in independent reactions to amplify segments of the mitochondrial 16S rRNA gene (L1865/H2195, hereafter Ac16s), 12S rRNA gene (L909/H1155, hereafter Am12s) and Cytochrome B gene (L14912/H15149c, hereafter CytB) (Evans et al. 2016, 2017; Olds et al. 2016). Amplified gene fragments were prepared for Illumina sequencing following a two-step PCR-based approach as outlined in the Illumina 16S Metagenomic Sequencing Library Preparation guidelines (Illumina, Inc., San Diego, CA, USA).

Amplified DNA from the first PCR were electrophoresed through a 2% agarose gel, stained with ethidium bromide and then visualised on a UV light platform. Each amplified product was manually excised from the gels using single-use razor blades, cleaned with a QIAquick Gel Extraction Kit (Qiagen, Venlo, Netherlands) and eluted from spin columns with 30µl of Buffer EB (Qiagen, Venlo, Netherlands). Regardless of visual confirmation of amplification, for both the mock communities and PCR negative control, we excised a band from the agarose gel at the expected amplicon size. For each sample, the gel extracted PCR products from the three gene regions were combined in proportions determined from previous work with these markers to approximate equal read production across genes (Am12S = 2.5 ng; Ac16S = 3.75 ng; CytB = 18.75 ng). In a second PCR, the Illumina adaptor sequence was added, including individualised library barcodes per sample. The DNA concentration of each library was quantified via Qubit dsDNA HS Assay. Libraries were pooled in equal molar concentrations along with 25% PhiX DNA (v3, Illumina, San Diego, California, USA) and then 300-bp paired-end sequenced on an Illumina MiSeq in two MiSeq flow cells in the University of Notre Dame's Genomics and Bioinformatics Core Facility (http://genomics.nd.edu/) using a MiSeq Reagent Kit v3 (600-cycle; Illumina, San Diego, California, USA).

Quality filtering, OTU clustering and species assignment

Bioinformatics analysis followed Olds et al. (2016) (scripts available on GitHub: https://github.com/pfrender-laboratory/epps). Briefly, raw sequence reads were removed from subsequent analysis if the read matched MiSeq sequencing adaptors > 6-bp, read quality score was smaller than 20 with a sliding window = 10bp or read length was shorter than 50-bp using Trimmomatic

version 0.32 (Bolger et al. 2014). Filtered paired-ended sequences were merged with USEARCH version 8.0.1623 (Edgar 2010). Identical merged reads were collapsed into one unique read using a custom Perl script (scripts available on GitHub: https://github.com/pfrender-laboratory/epps). The number of reads collapsed in unique reads was recorded. Unique reads were then clustered at 97% similarity with USEARCH version 8.0.1623 (Edgar 2010) to establish Operational Taxonomic Units (OTUs). The number of unique reads and their identical reads in each OTU was recorded and used as the abundance of the OTU. To accelerate the species assignment process, profile-hidden Markov models (HMMs) were used to remove sequences that are not of metazoan origin. A HMM model was constructed for each targeted gene region based on sequences of metazoan species from NCBI RefSeq Release 69 using HMMER version 3.1b1 (Wheeler and Eddy 2013). The centroid sequence of each OTU was mapped to the profile hidden Markov models to remove sequences that are not of metazoan origin. To establish the taxonomic affiliation of the OTUs, the centroid sequence of each OTU was assigned to species with SAP v1.9.4 (Munch et al. 2008) using NCBI NR database (version of April 2016) based on 97% sequence similarity to references (-minidentity 0.97). OTUs of mammal or bird origin were removed.

Contamination controls

No observable bands (PCR products) were detected on agarose gels in the expected size ranges for the eight field negative controls. Therefore, the eight field negative controls were removed from further analysis. After species assignment, potential contamination was removed based on the threshold correction method (Hänfling et al. 2016; Evans et al. 2017). In brief, the relative read abundance was calculated for each species found in each genetic marker in the control samples (four mock communities and one PCR negative control). The relative read abundance then functioned as a minimum detection threshold for each marker and species. Species in the samples detected at a lower relative read abundance than this threshold were removed from subsequent analysis. The script for the threshold correction method can be found on Github (https://github.com/lyy005/threshold_correction_for_eDNA/). Mock communities 1 and 2 and the PCR negative controls were used to estimate the contamination thresholds for samples from Locations 1-4. Mock communities 3 and 4 and the PCR negative controls were used to set the contamination thresholds for samples from Locations 5-8.

Comparisons between different genetic markers

To test whether the biodiversity patterns varied significantly between gene regions, an unfiltered and a filtered dataset were used for each region. A species (OTU) was considered detected if its DNA was found in any sample replicate or any marker from a site. This single occurrence criterion was used in the comparison of different gene regions since the "at least two markers" criterion in the moderate bioinformatics stringency (Evans et al. 2017) for species detection (see below) was not applicable for single gene regions. The unfiltered datasets used all the OTUs, including those not identified to species. In contrast, the filtered datasets included only the OTUs that could be identified to species-level. By this definition, inferences of species diversity will always be more conservative for the filtered than the unfiltered datasets because the filtered dataset is always a subset of the total OTUs detected.

Presence of species based on multiple genetic markers and replicates

Based on the sampling effort (three replicates for each site), the potential biodiversity in a relatively small environment and detection thresholds from previous studies, we utilised a moderate bioinformatics stringency (Evans et al. 2017) for species detection, namely a species was considered detected if its DNA was found in a least two replicate samples from a site or two markers from a single sample at a site.

Analyses of α- and β-diversity of samples along Eagle Creek

All analyses of α- and β-diversity were conducted in R version 3.3.2 (R Core Team 2016). The R-scripts can be found in figshare (https://doi.org/10.6084/m9.figshare.5860851). Genetic data were treated as incidence data. Alpha-diversity was calculated as species richness (number of OTUs assigned to species level) at each sampling location. Beta-diversity was estimated as Jaccard distance using the "vegdist" command in the "vegan" package version 2.4-2 of R (Oksanen et al. 2013). PCoA plots were made based on the β-diversity matrix with "cmdscale" command in R to visualise the relationships amongst replicate samples and sites.

The correlation between β-diversity matrices and the longitudinal distance matrix was tested with a Mantel test using "mantel.rtest" command in the R package "ade4" version 1.7-6 (Dray et al. 2007). Since Location 7 in the Kalamazoo River was upstream of the Eagle Creek confluence, it was not included in the correlation analysis between biodiversity and linear stream distance.

To evaluate the biodiversity detected by each replicate, the β-diversity between replicates was also calculated with "vegdist" based on Jaccard distance. Differences in β-diversities within and between the sample locations were tested using the Multi-Response Permutation Procedure (MRPP) in the R package "vegan" version 2.4-2 (Oksanen et al. 2013) using Jaccard distance.

Results

Environmental variables at each sampling location in Eagle Creek

The physical properties, including discharge and longitudinal distance from the headwaters of each sampling location along Eagle Creek are summarised in Suppl. materials 2 and 3, respectively. In general, discharge increased with distance downstream as expected. The headwaters of Eagle Creek showed the lowest discharge (Location 1 Q = 0.10 L/s) whereas the most downstream site in Eagle Creek had much higher discharge (Location 6 Q = 324 L/s). The two locations (Location 7 and 8) in the Kalamazoo River had the highest discharge (Q = 898 L/s).

High-throughput sequencing statistics and contamination controls

Illumina MiSeq sequencing generated 20.9 million raw reads. After filtering, merging and quality control, 2.1 million, 3.9 million and 3.0 million reads were used for downstream analysis for Ac16S, Am12S and CytB markers, across the total dataset (Table 1). After clustering, 54, 93 and 36 OTUs were detected with Ac16S, Am12S and Cytb markers, respectively. In these OTUs, 43, 79 and 36 OTUs were identified as vertebrates with HMM filtering. After the threshold correction, reads from mock community species were detected in the Am12S region of replicates 1a and 1b in Location 1 (Table 1). Therefore, Am12S marker of replicates 1a and 1b were removed from subsequent analysis. No other laboratory contamination was found. After the data filtering steps, 29, 33 and 15 OTUs were identified as fish species for Ac16S, Am12S and CytB markers, respectively (Table 1).

The identifications of fish species were refined based on their known geographic distributions. For example, the Ac16S and CytB markers were not diagnostic between *Umbra limi* and *Umbra pygmaea*. Since only *U. limi* is known from Michigan, all OTUs with an assignment to an *Umbra* spp. were assigned to *Umbra limi*. Similarly, the three gene regions were not diagnostic for the two closely related species, *Ambloplites ariommus* and *Ambloplites rupestris*. As only *A. rupestri* is found in Michigan, *Ambloplites* spp. was assigned to *A. rupestri*. The Am12S marker could also not distinguish between *Catostomus commersonii* and *Catostomus catostomus*. Since both are found in Michigan, the identification of these species was grouped at the generic level as *Catostomus* spp. *Notropis topeka* were removed as these species are not found in Michigan or nearby regions (NatureServe 2014).

α- and β-diversity of different genetic markers

Three gene regions were used for estimating α-diversity. When including all the OTUs (i.e. unfiltered dataset), the Am12S region detected the highest number of OTUs. When including only species-level assignments (i.e. filtered dataset), Ac16S detected the highest number of OTUs (Table 1). Only four species were detected by all three gene regions (Figure 2). The number of species co-detected by two markers was the highest for the pair Ac16S and Am12S (10 species). The pair Ac16S and CytB co-detected seven fish species, while Am12S and CytB had the lowest co-detection of species at four.

Different gene regions showed similar β-diversity patterns. For both the filtered and unfiltered data sets, β-diversity estimates, based on single gene regions and three markers, were positively correlated with each other. All estimates of β-diversity and longitudinal distance were positively correlated except between the Ac16S-filtered estimates and distance (all the *r* statistics and *p*-values are

Table 1. Summary of read number for each primer during the data analysis.

	Ac16s	Am12s	Cytb
Number of reads after quality control	2,137,148	3,861,079	3,010,080
Number of unique reads	208,072	131,601	201,606
Number of unique reads without singletons	53,645	46,396	43,363
Number of OTUs	54	93	36
Number of OTUs after HMM filtering	43	79	36
Number of OTUs after contamination removal (unfiltered dataset)	29	33	15
Number of OTUs with species-level assignment (filtered dataset)	20	15	9
Number of species (all three genetic markers combined)	23		

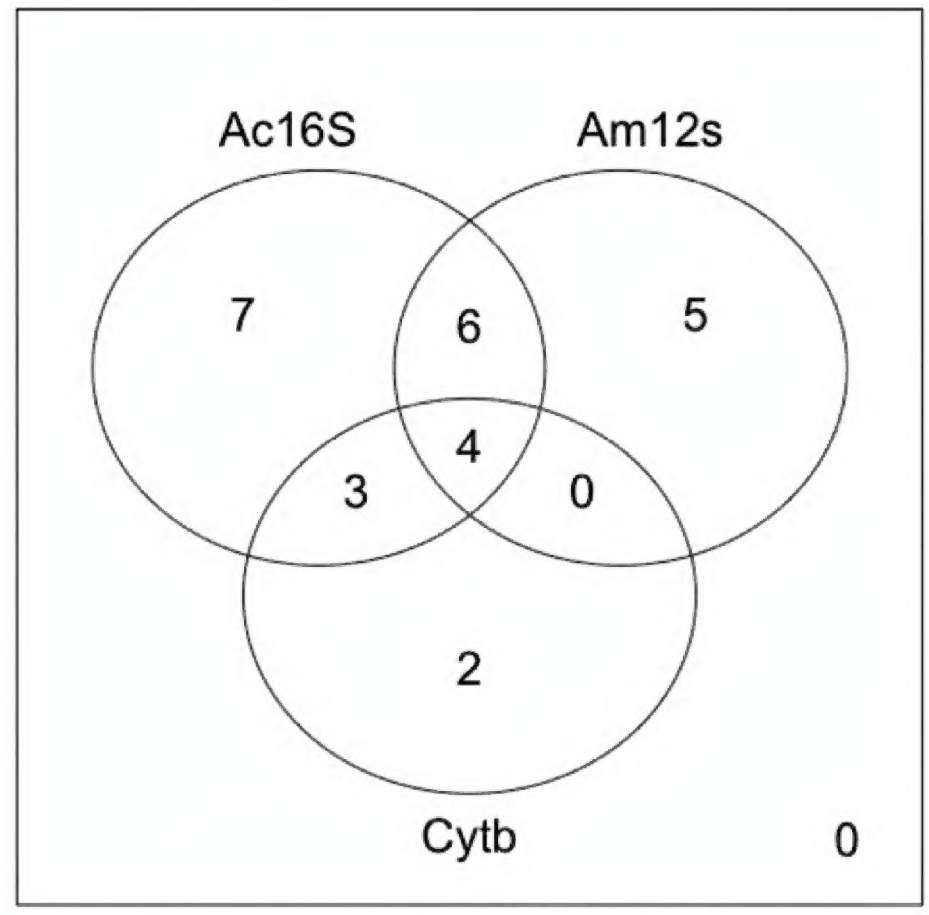

Figure 2. Venn diagram of number of species detected by each genetic marker based on filtered datasets.

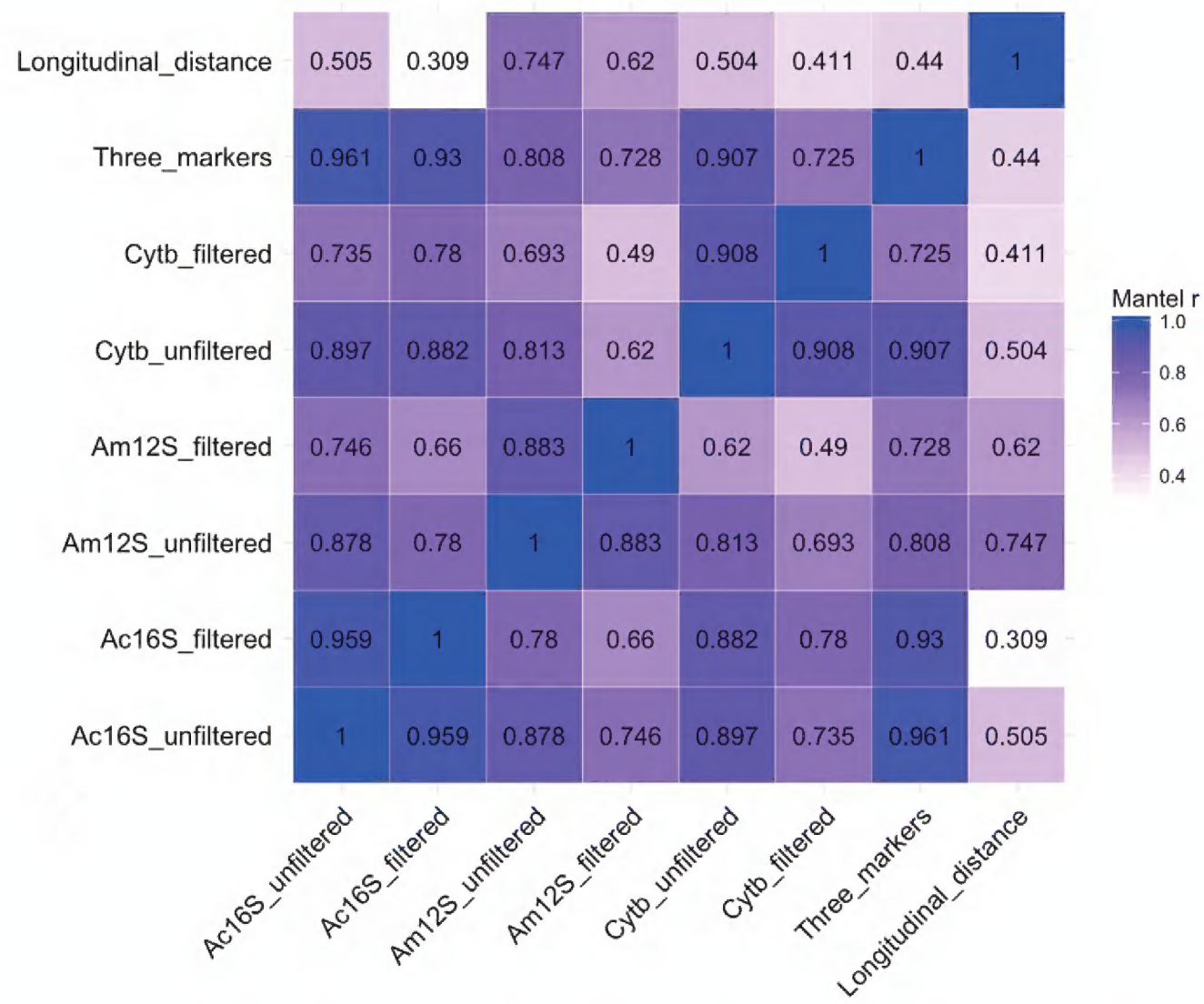

Figure 3. Heat map of correlations (Mantel *r*) between β-diversities as determined by the different genetic markers as well as the physical distance between samples. "Three_markers" refers to the combination of three filtered datasets (Ac16S_filtered, Am12S_filtered and Cytb_filtered).

Table 2. Fish species presence along Eagle Creek and in the Kalamazoo River detected using three genetic markers (Ac16S, Am12S and Cytb). Each column is a sampling location. Each row is the presence (x) or absence of each species.

	Sampling locations along Eagle Creek (1-6) and Kalamazoo River (7-8)							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Ambloplites rupestris				x	x		x	x
Ameiurus natalis		x	x	x	x	x	x	
Amia calva							x	x
Cyprinella spiloptera							x	
Erimyzon sucetta					x			
Esox americanus	x							
Etheostoma caeruleum					x	x		x
Etheostoma exile		x		x				
Etheostoma nigrum					x	x	x	x
Hypentelium nigricans							x	x
Lepomis cyanellus						x		
Lepomis gibbosus		x	x	x	x	x		
Lepomis gulosus						x		
Lepomis macrochirus		x	x	x	x	x	x	x
Micropterus dolomieu								x
Micropterus salmoides		x		x		x		x
Moxostoma erythrurum						x	x	x
Notemigonus crysoleucas		x	x					
Noturus flavus								x
Perca flavescens		x	x					
Percina maculata								x
Semotilus atromaculatus				x	x	x	x	
Umbra limi	x			x				
Total number of species	2	7	5	8	8	10	9	11

available in Suppl. material 4). The β-diversity estimates generated from the three unfiltered datasets (Ac16S unfiltered, Am12S unfiltered and CytB unfiltered) and three filtered datasets (Ac16S filtered, Am12S filtered and CytB filtered) were significantly correlated with each other. For each marker, the unfiltered datasets had a better correlation with the combined three marker dataset than with the corresponding single marker filtered dataset (Figure 3).

α-diversity in Eagle Creek

Based on the combined signal of all three genetic markers, 23 fish species were considered detected in the 6 sampling locations in Eagle Creek and 2 sampling locations in Kalamazoo River (Table 2). From upstream in Eagle Creek (Location 1) to downstream in the Kalamazoo River (Location 8), the number of fish species at a sampling site increased from 2 species to 11 species (Table 2). Location 1 had the lowest fish diversity (two species). Locations 2 and 3 had an intermediate fish diversity with seven and five species, respectively. Two species (*Notemigonus crysoleucas* and *Perca flavescens*) were only found in these two locations in Eagle Creek. Locations 4, 5 and 6 showed greater fish diversity with 8, 8 and 10 species detected, eight of which were shared between at least two of three locations. Furthermore, *Erimyzon sucetta* was only found in Location 5 and *Lepomis gulosus* was only found in Location 6. The two sampling locations in the Kalamazoo River (Locations 7 and 8) showed the highest number of species detected (9 and 11 species detected).

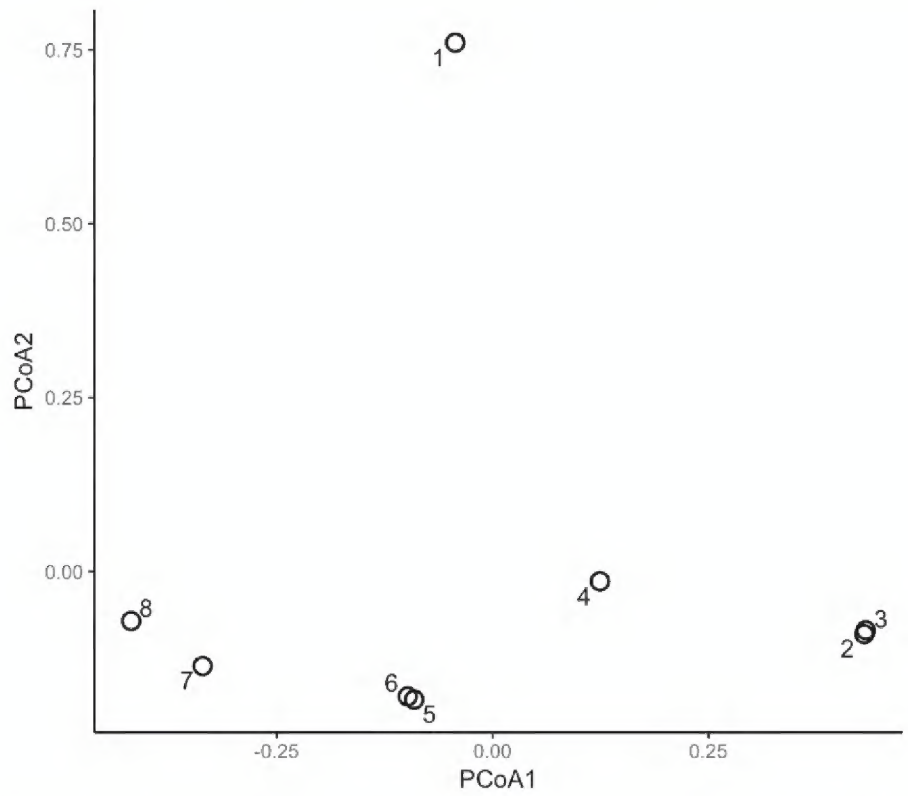

Figure 4. PCoA of the β-diversity of sampling locations with Jaccard distance based on all three genetic markers.

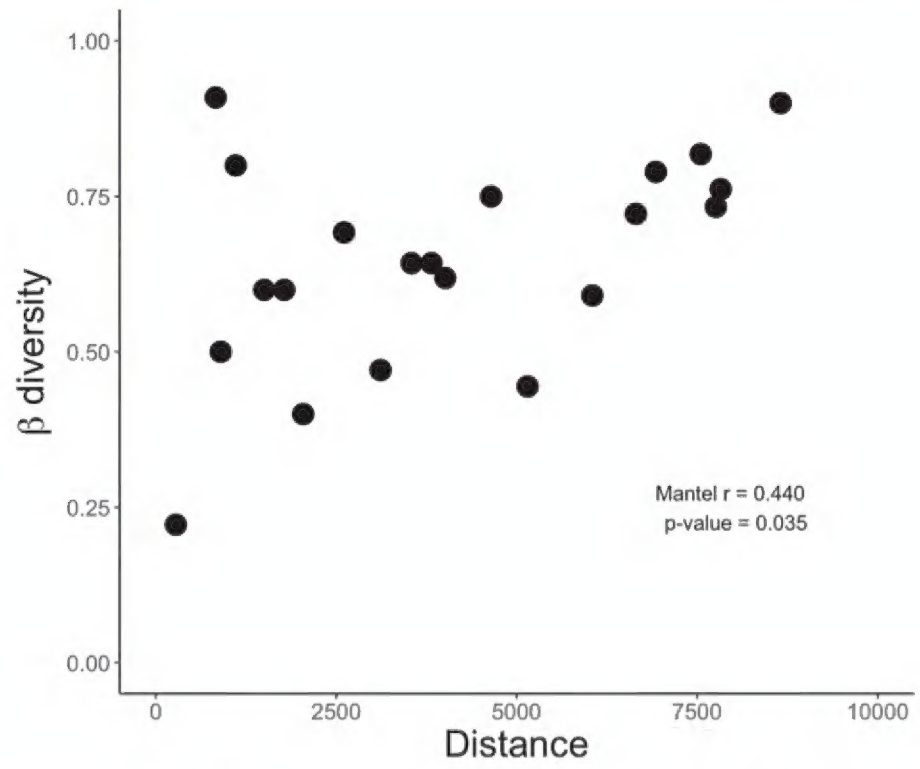

Figure 5. Relationship between the distance between sampling locations and β-diversity based on the combined results of all three genetic markers.

Noturus flavus and *Percina maculata* were only detected at Location 8, while *Cyprinella spiloptera* was only detected at Location 7.

β-diversity from upstream to downstream along Eagle Creek

Based on PCoA (Figure 4), the most upstream Location 1 had a distinct fish assemblage from other locations. Locations 2 and 3 displayed similar biodiversity, as did Locations 5 and 6 and Locations 7 and 8. Fish diversity in Location 4 was intermediate between Locations 3 and 5. The longitudinal stream distance between sampling locations and β-diversity were positively correlated (Figure 5). In addition to distance, other environmental variables (temperature, pH, turbidity) significantly varied with β-diversity (Suppl. material 5). These variables (temperature, pH, D.O., discharge) are not independent of distance (Suppl. material 5) or other environmental variables. Based on the β-diversity measures amongst all the replicates, replicates from the same location were more similar than replicates between locations with a MRPP test (p-value < 0.01) (the PCoA plot and table with species and their read number in each replicate can be found in Suppl. materials 6 and 7).

Discussion

Fish species diversity in Eagle Creek

In this study, we used eDNA metabarcoding to describe fish diversity along the course of a small stream. The local fish assemblage varied spatially from upstream to downstream, consistent with previous studies (Magalhaes et al. 2002; Vannote et al. 1980). In stream headwaters, fish species diversity tends to be low relative to downstream sites where habitat diversity increases and species tend to accumulate along the river continuum (Tornwall et al. 2015). The most upstream headwater site of Eagle Creek had a low-diversity and distinctive fish assemblage. The two locations in the much larger Kalamazoo River had the highest number of fish species, consistent with predictions of the River Continuum Concept (Vannote et al. 1980).

Fish species assemblages also showed absence, additions and turnovers (as measured using β-diversity) along the stream gradient. Samples collected from adjacent locations often showed similar species compositions and dissimilarity increased with distance between sites. The differences in species composition along the stream could be due to changes in habitat structure and availability, as well as changes in temperature, pH and dissolved oxygen along the stream course (Almeida and Cetra 2016). The two sampling locations (Locations 7 and 8) upstream and downstream of the confluence of Eagle Creek, respectively, showed different fish species composition. Eagle Creek likely contributed eDNA to the downstream site of the Kalamazoo River. For example, *Etheostoma caeruleum* and *Micropterus salmoides* were detected in Eagle Creek and at Location 8 in the Kalamazoo River where water from both systems had mixed but not at Location 7 upstream of the confluence. The absence of some fish species at Location 8 that were detected at Location 7 may be related to the distance between sites or dilution by incoming water from Eagle Creek.

Replicates and multiple markers increase the reliability of eDNA diversity detection

We collected three replicate eDNA samples in each location to improve species inference following our practice (Olds et al. 2016; Evans et al. 2017) and as suggested by Goldberg et al. (2016). Based on a MRPP test and the PCoA analysis, replicates collected from the same location

displayed similar biodiversity, although some variation was detected. We collected a modest number of replicates (3 replicates of 250 ml water) for each sampling location. Although more replicates could improve the detection of low-abundance species (Evans et al. 2017), the low variance amongst the replicates indicates that additional replicates may have had a minor effect on our results. We also found a positive correlation between β-diversity amongst sample sites and stream distance, indicating our site sampling scheme generated sufficient statistical power to infer the patterns of fish community similarity within Eagle Creek. Our previous studies suggested that use of a moderate bioinformatics stringency ("at least two markers or two samples" criterion) can minimise the effect of primer bias (Evans et al. 2016), minimise cross-sample contamination (Olds et al. 2016) and improve the reliability of species identification (Evans et al. 2017). In this study, we found that the overall biodiversity patterns (β-diversity estimates) inferred by the different genetic markers were similar. This finding is consistent with previous research that β-diversity estimates are less sensitive to the specific genetic markers employed than β-diversity estimates (Drummond et al. 2015). By comparing the filtered and unfiltered dataset of each genetic marker to the three-marker dataset, we found that including OTUs that did not have a species-level assignment (unfiltered datasets) added more information about community similarities amongst samples. In general, our results suggest that different genetic markers express the same β-diversity pattern along the longitudinal stream gradient in Eagle Creek.

eDNA transport

Understanding the spatial distribution of eDNA in flowing environments is critical for accurate detection of species in stream ecosystems (Deiner et al. 2016; Jerde et al. 2016; Shogren et al. 2017). Environmental DNA transported from upstream may lead to false positive detection of species in downstream locations and it becomes difficult to infer species location relative to sampling location.

Based on previous studies, the average transport distance of eDNA can be highly variable, depending on stream discharge (Barnes and Turner 2016; Jerde et al. 2016; Shogren et al. 2016, 2017; Wilcox et al. 2016). For example, studies in low discharge (Q = 2-10 L/s) experimental streams found that eDNA transport distance was 50-1459 m (Jerde et al. 2016; Wilcox et al. 2016), but eDNA has been detected up to 10 km downstream in high discharge systems (Q > 3,000 L/s, Deiner and Altermatt 2014). In this study, we estimated β-diversity between sampling sites and showed that these estimates were correlated with stream distance. The implication of this result is that the downstream transport of eDNA in this low discharge stream is insufficient to completely homogenise the signal of fish diversity at individual sites. The low discharge in the headwaters of Eagle Creek likely limited the eDNA transport distance to less than the between site distances. For example, Locations 1 and 2 had no species overlap even though they were separated by only 829 m. These differences in species composition were probably not the result of low abundance and low detection probabilities since the most abundant species (based on sequence read number) in Location 1, *Umbra limi*, was not detected at Location 2. Given these results and the low discharges measured at Location 1 and 2 (0.1 L/s and 73.5 L/s respectively), it is likely that the DNA transport distance is less than 829 m in the headwaters of Eagle Creek. Sampling eDNA in low discharge streams is potentially a useful way for managers to measure diversity for more local scales. Higher-discharge downstream locations in Eagle Creek shared species including most of the high-abundance species (Table 2) so we cannot distinguish whether the detected species' eDNA at any particular locality was derived from eDNA transport or the mobility of species (Pilliod et al. 2013).

Our goal was not to mechanistically describe why β-diversity changes in the river, but to test whether transport of eDNA obscures the signal of species turnover in lower discharge streams. The influence that environmental variables (turbidity, D.O., temperature etc.) have on the transport and detection of eDNA is not well understood, although, increasingly, studies are finding evidence that environmental variables can affect how long eDNA remains detectable in a system (Seymour et al. 2018; Shogren et al. 2017). Additional mechanistic tests of such variables are needed to understand when and how the environmental conditions in streams influence the detection and transport of eDNA (Deiner et al. 2017). By understanding the physical and ecological mechanisms of eDNA behaviour in lotic environments, we can better estimate the presence of species in rivers and streams using eDNA metabarcoding.

Conclusions

In this study, we estimated the longitudinal fish biodiversity within a stream using a multi-marker eDNA metabarcoding approach. We found that the α-diversity of Eagle Creek increased from upstream to downstream and that β-diversity was positively correlated with the longitudinal distance between locations. We also found that the observed biodiversity patterns were similar amongst different genetic markers and replicate samples. Our results suggest that eDNA can be transported hundreds of metres in a small stream, consistent with previous studies. Additional studies are needed to better quantify and understand eDNA transport dynamics in streams of varying discharge.

Data accessibility

Raw sequence data is available on the NCBI repository under the BioProject PRJNA317862 with accession numbers: SRX3626969–SRX3626997; OTU sequences and R scripts used in this study are available on figshare (https://doi.org/10.6084/m9.figshare.5860851).

Authors' contributions

Sampling: NTE, BPO;
Laboratory work: MAR;
Data analyses: YL, NTE, MAR, CLJ, BPO, AJS, KD;
Study design and supervision: DML, GAL, MEP;
Manuscript writing: all authors edited and commented on the manuscript draft.

Acknowledgements

Funding for this research was provided by United States Department of Defense Strategic Environmental Research and Development Program Grant W912Hq-12-C-0073 (RC-2240). We thank Stuart E. Jones for help with statistical analyses, Crysta A. Gantz for assistance with field sampling and eDNA sample processing, Cameron R. Turner for help with sampling and experimental design and Jacqueline Lopez in the Notre Dame Genomics and Bioinformatics Core Facility for assistance with library QC and sequencing. We thank Michele Richards and other personnel of Fort Custer Training Center (Michigan Army National Guard) for site access and logistical support. We thank David C. Molik (University of Notre Dame), Paul Czechowski (Cornell University) for advice and helpful discussions.

References

Almeida RS, Cetra M (2016) Longitudinal gradient effects on the stream fish metacommunity. Natureza & Conservação 14: 112–119. https://doi.org/10.1016/j.ncon.2016.10.001

Bae M-J, Li F, Kwon Y-S, Chung N, Choi H, Hwang S-J, Park Y-S (2014) Concordance of diatom, macroinvertebrate and fish assemblages in streams at nested spatial scales: Implications for ecological integrity. Ecological Indicators 47: 89–101. https://doi.org/10.1016/j.ecolind.2014.07.030

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics 17: 1–17. https://doi.org/10.1007/s10592-015-0775-4

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Douglas WY, De Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology & Evolution 29: 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics btu170. https://doi.org/10.1093/bioinformatics/btu170

Bonar SA, Hubert WA, Willis DW (2009) Standard methods for sampling North American freshwater fishes.

Civade R, Dejean T, Valentini A, Roset N, Raymond J-C, Bonin A, Taberlet P, Pont D (2016) Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. PLoS One 11: e0157366. https://doi.org/10.1371/journal.pone.0157366

Deiner K, Altermatt F (2014) Transport distance of invertebrate environmental DNA in a natural river. PLoS One 9: e88786. https://doi.org/10.1371/journal.pone.0088786

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, others (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. Molecular Ecology. https://doi.org/10.1111/mec.14350

Deiner K, Fronhofer EA, Mächler E, Walser J-C, Altermatt F (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature communications 7. https://doi.org/10.1038/ncomms12544

Dray S, Dufour A-B, others (2007) The ade4 package: implementing the duality diagram for ecologists. Journal of statistical software 22: 1–20. https://doi.org/10.18637/jss.v022.i04

Drummond AJ, Newcomb RD, Buckley TR, Xie D, Dopheide A, Potter BC, Heled J, Ross HA, Tooman L, Grosser S, others (2015) Evaluating a multigene environmental DNA approach for biodiversity assessment. GigaScience 4: 46. https://doi.org/10.1186/s13742-015-0086-1

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Evans NT, Lamberti GA (2017) Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool. Fisheries Research. https://doi.org/10.1139/cjfas-2016-0306

Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL, Lodge DM, Lamberti GA, Pfrender ME (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. Canadian Journal of Fisheries and Aquatic Sciences.

Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular ecology resources 16: 29–41. https://doi.org/10.1111/1755-0998.12433

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biology letters 4: 423–425. https://doi.org/10.1098/rsbl.2008.0118

Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, Spear SF, McKee A, Oyler-McCance SJ, Cornman RS, others (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution 7: 1299–1307. https://doi.org/10.1111/2041-210X.12595

Gore JA, Banning J (2017) Discharge measurements and streamflow analysis. Methods in Stream Ecology, Volume 1 (Third Edition). Elsevier, 49–70.

Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular ecology 25: 3101–3119. https://doi.org/10.1111/mec.13660

Holmlund CM, Hammer M (1999) Ecosystem services generated by fish populations. Ecological economics 29: 253–268. https://doi.org/10.1016/S0921-8009(99)00015-4

Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. Conservation Letters 4: 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jerde CL, Olds BP, Shogren AJ, Andruszkiewicz EA, Mahon AR, Bolster D, Tank JL (2016) Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. Environmental Science & Technology 50: 8770–8779. https://doi.org/10.1021/acs.est.6b01761

Lodge DM, Turner CR, Jerde CL, Barnes MA, Chadderton L, Egan SP, Feder JL, Mahon AR, Pfrender ME (2012) Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. Molecular Ecology 21: 2555–2558. https://doi.org/10.1111/j.1365-294X.2012.05600.x

Magalhaes MF, Batalha DC, Collares-Pereira MJ (2002) Gradients in stream fish assemblages across a Mediterranean landscape: contributions of environmental factors and spatial structure. Freshwater Biology 47: 1015–1031. https://doi.org/10.1046/j.1365-2427.2002.00830.x

Marzin A, Archaimbault V, Belliard J, Chauvin C, Delmas F, Pont D (2012) Ecological assessment of running waters: do macrophytes, macroinvertebrates, diatoms and fish show similar responses to human pressures? Ecological indicators 23: 56–65. https://doi.org/10.1016/j.ecolind.2012.03.010

McInerney PJ, Rees GN (2018) More (or less?) bounce for the ounce: a comparison of environmental DNA and classical approaches for bioassessment. Marine and Freshwater Research. https://doi.org/10.1071/MF17250

McIntyre PB, Liermann CAR, Revenga C (2016) Linking freshwater fishery management to global food security and biodiversity conservation. Proceedings of the National Academy of Sciences 113: 12880–12885. https://doi.org/10.1073/pnas.1521540113

Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008) Statistical assignment of DNA sequences using Bayesian phylogenetics. Systematic biology 57: 750–757. https://doi.org/10.1080/10635150802422316

Nature Serve (2014) Notropis tepeka. The IUCN Red List of Threatened Species 2014: e.T184092A19034321. http://dx.doi.org/10.2305/IUCN.UK.2014-3.RLTS.T184092A19034321.en

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H, others (2013) Package "vegan." Community ecology package, version 2.

Olds BP, Jerde CL, Renshaw MA, Li Y, Evans NT, Turner CR, Deiner K, Mahon AR, Brueseke MA, Shirey PD, others (2016) Estimating species richness using environmental DNA. Ecology and Evolution 6: 4214–4226. https://doi.org/10.1002/ece3.2186

Pilliod DS, Goldberg CS, Arkle RS, Waits LP (2014) Factors influencing detection of eDNA from a stream-dwelling amphibian. Molecular Ecology Resources 14: 109–116. https://doi.org/10.1111/1755-0998.12159

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Renshaw MA, Olds BP, Jerde CL, McVeigh MM, Lodge DM (2015) The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. Molecular ecology resources 15: 168–176. https://doi.org/10.1111/1755-0998.12281

Seymour M, Durance I, Cosby BJ, Ransom-Jones E, Deiner K, Ormerod SJ, Colbourne JK, Wilgar G, Carvalho GR, de Bruyn M, others (2018) Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. Communications Biology 1: 4. https://doi.org/10.1038/s42003-017-0005-3

Shogren AJ, Tank JL, Andruszkiewicz E, Olds B, Mahon AR, Jerde CL, Bolster D (2017) Controls on eDNA movement in streams: Transport, Retention, and Resuspension. Scientific Reports 7. https://doi.org/10.1038/s41598-017-05223-1

Shogren AJ, Tank JL, Andruszkiewicz EA, Olds B, Jerde C, Bolster D (2016) Modelling the transport of environmental DNA through a porous substrate using continuous flow-through column experiments. Journal of The Royal Society Interface 13: 20160290. https://doi.org/10.1098/rsif.2016.0290

Simon TP, Evans NT (2017) Environmental Quality Assessment Using Stream Fishes. Methods in Stream Ecology (Third Edition). Elsevier, 319–334. https://doi.org/10.1016/B978-0-12-813047-6.00017-6

Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Orlando L, Willerslev E (2012) Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology 21: 2565–2573. https://doi.org/10.1111/j.1365-294X.2011.05418.x

Tornwall B, Sokol E, Skelton J, Brown BL (2015) Trends in stream biodiversity research since the river continuum concept. Diversity 7: 16–35. https://doi.org/10.3390/d7010016

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE (1980) The river continuum concept. Canadian journal of fisheries and aquatic sciences 37: 130–137. https://doi.org/10.1139/f80-017

Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, Glidden S, Bunn SE, Sullivan CA, Liermann CR, others (2010) Global threats to human water security and river biodiversity. Nature 467: 555. https://doi.org/10.1038/nature09440

Wheeler TJ, Eddy SR (2013) nhmmer: DNA homology search with profile HMMs. Bioinformatics btt403. https://doi.org/10.1093/bioinformatics/btt403

Wilcox TM, McKelvey KS, Young MK, Sepulveda AJ, Shepard BB, Jane SF, Whiteley AR, Lowe WH, Schwartz MK (2016) Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. Biological Conservation 194: 209–216. https://doi.org/10.1016/j.biocon.2015.12.023

Supplementary material 1

Control samples that were used to test for contamination during each step of the experiment.

Authors: Yiyuan Li, Nathan T. Evans , Mark A. Renshaw, Christopher L. Jerde, Brett P. Olds, Arial J. Shogren, Kristy Deiner, David M. Lodge, Gary A. Lamberti, Michael E. Pfrender

Data type: (measurement/occurence/multimedia/etc.)

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.24262.suppl1

Supplementary material 2
GPS and environmental variables of each site.
Authors: Yiyuan Li, Nathan T. Evans , Mark A. Renshaw, Christopher L. Jerde, Brett P. Olds, Arial J. Shogren, Kristy Deiner, David M. Lodge, Gary A. Lamberti, Michael E. Pfrender
Data type: (measurement/occurence/multimedia/etc.)
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.24262.suppl2

Supplementary material 3
The longitudinal distance (upper triangular) and β- diversity (lower triangular) between sampling locations along Eagle Creek.
Authors: Yiyuan Li, Nathan T. Evans , Mark A. Renshaw, Christopher L. Jerde, Brett P. Olds, Arial J. Shogren, Kristy Deiner, David M. Lodge, Gary A. Lamberti, Michael E. Pfrender
Data type: (measurement/occurence/multimedia/etc.)
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.24262.suppl3

Supplementary material 4
Mantel r and p-values for all the pairwise comparisons between single marker, three markers and longitudinal distance.
Authors: Yiyuan Li, Nathan T. Evans , Mark A. Renshaw, Christopher L. Jerde, Brett P. Olds, Arial J. Shogren, Kristy Deiner, David M. Lodge, Gary A. Lamberti, Michael E. Pfrender
Data type: (measurement/occurence/multimedia/etc.)
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.24262.suppl4

Supplementary material 5
The correlation between environmental variables and β-diversity and longitudinal distance using Mantel test
Authors: Yiyuan Li, Nathan T. Evans , Mark A. Renshaw, Christopher L. Jerde, Brett P. Olds, Arial J. Shogren, Kristy Deiner, David M. Lodge, Gary A. Lamberti, Michael E. Pfrender
Data type: (measurement/occurence/multimedia/etc.)
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.24262.suppl5

Supplementary material 6
PCoA plot based on all the replicates from Eagle Creek
Authors: Yiyuan Li, Nathan T. Evans , Mark A. Renshaw, Christopher L. Jerde, Brett P. Olds, Arial J. Shogren, Kristy Deiner, David M. Lodge, Gary A. Lamberti, Michael E. Pfrender
Data type: (measurement/occurence/multimedia/etc.)
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.24262.suppl6

Supplementary material 7
Profiling table for all the replicates.
Authors: Yiyuan Li, Nathan T. Evans , Mark A. Renshaw, Christopher L. Jerde, Brett P. Olds, Arial J. Shogren, Kristy Deiner, David M. Lodge, Gary A. Lamberti, Michael E. Pfrender
Data type: (measurement/occurence/multimedia/etc.)
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.24262.suppl7